

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2001, 18:12:04 ; Search time 18.37 Seconds

(without alignments)
1156.146 Million cell updates/sec

Title: US-09-405-504A-49

Perfect score: 3271

Sequence: 1 MSAITVTLGLFLPLLVN.....MYVPMTEDLYNAISAKTLK 620

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3271	100.0	620	1 VLCS_HUMAN	O14975 homo sapien
2	2801	85.6	620	1 VLCS_RAT	P97524 ratius norv
3	2798	85.5	620	1 VLCS_MOUSE	O35488 mus musculu
4	1044	31.9	646	1 FATP_MOUSE	O60714 mus musculu
5	1042.5	31.9	646	1 FATP_RAT	P97849 ratius norv
6	744.5	22.8	623	1 FATL_YEAST	P38225 saccharomyc
7	398.5	12.2	522	1 CAIC_ECOLI	P13532 escherichia
8	276.5	8.3	561	1 LCFA_ECOLI	P29212 escherichia
9	274.5	8.4	553	1 ACL_VANPL	O24540 vanilla pla
10	272.5	8.3	556	1 ACL2_ARATH	O95725 arabidopsis
11	270.5	8.3	563	1 ACL1_ORYSA	P17814 oryza sativ
12	266.5	8.1	544	1 SREL_BACSU	P27206 bacillus su
13	265.5	8.1	544	1 ACL1_PENCR	P14912 petroselinu
14	264.5	8.1	562	1 LCFA_HAERIN	P46450 haemophilus
15	261.5	8.0	537	1 ACL_PINTA	P41636 pinus taeda
16	261.5	8.0	544	1 ACL2_PENCR	P14913 petroselinu
17	260.5	8.0	547	1 ACL1_TOBAC	O24145 nicotiana t
18	259.5	7.9	1274	1 SREL_BACSU	O08787 bacillus su
19	258.5	7.9	561	1 PPS2_BACSU	P39846 bacillus su
20	257.5	7.9	561	1 ACL3_ARATH	O95777 arabidopsis
21	255.5	7.8	542	1 ACL2_TOBAC	O24146 nicotiana t
22	255.5	7.8	545	1 ACL2_SOLTU	P31684 solanum tub
23	254.5	7.8	545	1 ACL1_SOLTU	P31685 solanum tub
24	251.5	7.7	6359	1 BACC_BACLI	O68008 b bacitract
25	250	7.6	2555	1 PPS3_BACSU	P39847 bacillus su
26	247.5	7.5	4451	1 GRSB_BACBR	P14688 b gramlicidi
27	246.5	7.5	543	1 FAT2_YEAST	P38137 saccharomyc
28	243	7.4	561	1 ACL1_ARATH	O42524 arabidopsis
29	238.5	7.3	670	1 ACSEA_EMBRI	P16928 emericeella
30	238.5	7.3	6486	1 TYCC_BACBR	O30409 b tyrocidin
31	237.5	7.3	5255	1 BACA_BACLI	O68006 b bacitract
32	236	7.2	3587	1 SRF2_BACSU	O04747 bacillus su
33	235.5	7.2	572	1 ACSEA_BACSU	P39062 bacillus su

34	233	7.1	560	1 LCFA_BACSU	P94547 bacillus su
35	229	7.0	550	1 LUCC_PROPY	P08659 photinus py
36	220	6.7	2607	1 BACC_BACLI	O68007 bacillus li
37	212	6.5	669	1 ACSEA_PENCR	P36333 penicillium
38	212	6.5	3587	1 TYCC_BACBR	O30408 b tyrocidin
39	211	6.5	653	1 ACSEA_SYNR3	O55404 synechocyst
40	210.5	6.4	672	1 ACSEA_METSO	P27095 methanocithr
41	210	6.4	662	1 ACSEA_HELPY	O25686 halicobacte
42	208.5	6.4	675	1 ACSEA_CANAL	O94049 candida alb
43	207	6.3	661	1 ACSEA_COPCI	O13440 penicillium ci
44	204.5	6.3	713	1 ACSEA_YEAST	O01574 saccharomyc
45	203.5	6.2	503	1 DITFA_BACSU	P39581 bacillus su

ALIGNMENTS

RESULT 1	VLCS_HUMAN	STANDARD:	PRT:	620 AA.
ID	VLCS_HUMAN			
AC	O14975;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DE	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	VERY-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.-) (VERY-LONG-CHAIN-FATTY-ACID-COA LIGASE).			
GN	FACVLI OR VLACS.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RA	Uchiyama A., Aoyama T., Kamiyo K., Wakui K., Fukushima Y.,			
RU	Shimozawa N., Suzuki Y., Kondo N., Orit T., Hashimoto T.;			
RN	Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99216327; PubMed=10198260;			
RA	Steinberg S.J., Wang S.J., Kim D.G., Mihalik S.J., Watkins P.A.;			
RT	"Human very-long-chain acyl-CoA synthetase: cloning, topography, and			
RT	relevance to branched-chain fatty acid metabolism."			
RL	Biochem. Biophys. Res. Commun. 257:615-621(1999).			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.			
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, PLACENTA AND			
CC	PANCREAS.			
CC	-1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT			
CC	COVALENT BINDING OF AMP TO THEIR SUBSTRATE.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/			
CC	or send an email to license@isb-sdb.ch).			
DR	EMBL; D88308; BAA23644.1; -			
DR	EMBL; AF096290; AAC64973.1; -			
DR	MIM; 603247; -			
DR	InterPro; IPR000873; -			
DR	Pfam; PF00501; AMP-binding; 1.			
DR	PROSITE; PS00455; AMP-BINDING; 1.			
KW	Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.			
FT	TRANSMEM	1	21	POTENTIAL.
FT	TRANSMEM	107	127	POTENTIAL.
FT	TRANSMEM	262	282	POTENTIAL.
SO	SEQUENCE	620 AA;	70312 MW;	BFD33E1BB67B3DF8 CAC64;

Query Match 100.0%; Score 3271; DB 1; Length 620;
Best Local Similarity 100.0%; Pred. No. 1e-239;

Matches 620; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSAITVYLAGLLEPLLVNLCCEPFODIGYFLKVAAGRVRSYGGRRPARTILRAFL 60
 DB 1 MSAITVYLAGLLEPLLVNLCCEPFODIGYFLKVAAGRVRSYGGRRPARTILRAFL 60
 QY 61 EKAROTPHKPELLEFDELTLYAQVDRRSNOVARALHDHGLRGDCVALMGNEPAYWL 120
 DB 61 EKAROTPHKPELLEFDELTLYAQVDRRSNOVARALHDHGLRGDCVALMGNEPAYWL 120
 QY 121 WGLVKGACMACLVNIRAKSLHCFQCCGAKVLLVSPDLAAVEELLPSLKDDVSIY 180
 DB 121 WGLVKGACMACLVNIRAKSLHCFQCCGAKVLLVSPDLAAVEELLPSLKDDVSIY 180
 QY 181 YVSRSTNDGIDISFLDKVDESTEPIPSWSEVTSTPALITYSGTGLPKAMITHQ 240
 DB 181 YVSRSTNDGIDISFLDKVDESTEPIPSWSEVTSTPALITYSGTGLPKAMITHQ 240
 QY 241 RIWGTGLTFVSGLRADVIYITLPEYHSAALLIGHCIYAGATLALRTKFSASQFMD 300
 DB 241 RIWGTGLTFVSGLRADVIYITLPEYHSAALLIGHCIYAGATLALRTKFSASQFMD 300
 QY 301 CRKYNTVYQYIGELLRYLCNSPQKPNDRDHVRLALGNGLRGQVWRQFVRRPDCIYE 360
 DB 301 CRKYNTVYQYIGELLRYLCNSPQKPNDRDHVRLALGNGLRGQVWRQFVRRPDCIYE 360
 QY 361 FFAATEGNIGFMNVARKVAVGVNVLQKITYDLIKYDEKDEPVDENGCYVVRPKG 420
 DB 361 FFAATEGNIGFMNVARKVAVGVNVLQKITYDLIKYDEKDEPVDENGCYVVRPKG 420
 QY 421 EYGLLVCKITTLTPENGVAAGAKOTEKKLRDVEFKKGLDLYNSGDLMAVDHENTFYFHDR 480
 DB 421 EYGLLVCKITTLTPENGVAAGAKOTEKKLRDVEFKKGLDLYNSGDLMAVDHENTFYFHDR 480
 QY 481 VGDTPFKMGENATTEVADTVGLVDFVQEVNNGVAVHPDHEGRIGMASIKKENHEFPKG 540
 DB 481 VGDTPFKMGENATTEVADTVGLVDFVQEVNNGVAVHPDHEGRIGMASIKKENHEFPKG 540
 QY 541 KLFQHIADYLPYSARPRFLRIODTIEITGTFKRRKMTLVEEGFNPAVAKDALYFLDDTAK 600
 DB 541 KLFQHIADYLPYSARPRFLRIODTIEITGTFKRRKMTLVEEGFNPAVAKDALYFLDDTAK 600
 QY 601 MYVPMEDIIYNAISAKTKL 620
 DB 601 MYVPMEDIIYNAISAKTKL 620

RESULT 2

QY VCLS_RAT STANDARD: PRT: 620 AA.
 ID VCLS_RAT
 AC P97524;
 DT 15-JUL-1998 (rel. 36, Created)
 DT 15-JUL-1998 (rel. 36, Last sequence update)
 DT 15-DEC-1998 (rel. 37, Last annotation update)
 DE VERR-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.1-) (VERY-LONG-CHAIN-FATTY-ACYL-COA LIGASE).
 GN FACYL OR VLACS.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 NC NCHL_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=WISTAR; TISSUE=Liver;
 RX MEDLINE=97094763; PubMed=8939997;
 RA Uchiyama A., Aoyama T., Kamijo K., Uchida Y., Kondo N., Orit T., Hashimoto T.;
 RT "Molecular cloning of cDNA encoding rat very long-chain acyl-CoA synthetase.";
 CC J. Biol. Chem. 271:30360-30365(1996).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL collaboration - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb.ch).

CC EMBL: D85100; BAA12722.1; -
 DR InterPro: IPR000873; -
 DR Pfam: PF00501; AMP-binding; 1.
 DR PROSITE: PS00455; AMP BINDING; 1.
 KW ligase; fatty acid metabolism; Peroxisome; Transmembrane.
 FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 107 127 POTENTIAL.
 FT TRANSMEM 262 282 POTENTIAL.
 SQ SEQUENCE 620 AA; 70693 MW; 6CF9362DC3805526 CRC64;

Query Match 85.6%; Score 2801; DB 1; Length 620;
 Best Local Similarity 82.4%; Pred. No. 3.5e-204;
 Matches 511; Conservative 55; Mismatches 54; Indels 0; Gaps 0;

QY 1 MSAITVYLAGLLEPLLVNLCCEPFODIGYFLKVAAGRVRSYGGRRPARTILRAFL 60
 DB 1 MSAITVYLAGLLEPLLVNLCCEPFODIGYFLKVAAGRVRSYGGRRPARTILRAFL 60
 QY 61 EKAROTPHKPELLEFDELTLYAQVDRRSNOVARALHDHGLRGDCVALMGNEPAYWL 120
 DB 61 EKAROTPHKPELLEFDELTLYAQVDRRSNOVARALHDHGLRGDCVALMGNEPAYWL 120
 QY 61 EQARKTTHKPELLEFDELTLYAQVDRRSNOVARALHDHGLRGDCVALMGNEPAYWL 120
 DB 61 EQARKTTHKPELLEFDELTLYAQVDRRSNOVARALHDHGLRGDCVALMGNEPAYWL 120
 QY 121 WGLVKGACMACLVNIRAKSLHCFQCCGAKVLLVSPDLAAVEELLPSLKDDVSIY 180
 DB 121 WGLVKGACMACLVNIRAKSLHCFQCCGAKVLLVSPDLAAVEELLPSLKDDVSIY 180
 QY 121 WGLVKGACMACLVNIRAKSLHCFQCCGAKVLLVSPDLAAVEELLPSLKDDVSIY 180
 DB 121 WGLVKGACMACLVNIRAKSLHCFQCCGAKVLLVSPDLAAVEELLPSLKDDVSIY 180
 QY 181 YVSRSTNDGIDISFLDKVDESTEPIPSWSEVTSTPALITYSGTGLPKAMITHQ 240
 DB 181 YVSRSTNDGIDISFLDKVDESTEPIPSWSEVTSTPALITYSGTGLPKAMITHQ 240
 QY 181 YVSRSTNDGIDISFLDKVDESTEPIPSWSEVTSTPALITYSGTGLPKAMITHQ 240
 DB 181 YVSRSTNDGIDISFLDKVDESTEPIPSWSEVTSTPALITYSGTGLPKAMITHQ 240
 QY 241 RIWGTGLTFVSGLRADVIYITLPEYHSAALLIGHCIYAGATLALRTKFSASQFMD 300
 DB 241 RIWGTGLTFVSGLRADVIYITLPEYHSAALLIGHCIYAGATLALRTKFSASQFMD 300
 QY 241 RIWGTGLTFVSGLRADVIYITLPEYHSAALLIGHCIYAGATLALRTKFSASQFMD 300
 DB 241 RIWGTGLTFVSGLRADVIYITLPEYHSAALLIGHCIYAGATLALRTKFSASQFMD 300
 QY 301 CRKYNTVYQYIGELLRYLCNSPQKPNDRDHVRLALGNGLRGQVWRQFVRRPDCIYE 360
 DB 301 CRKYNTVYQYIGELLRYLCNSPQKPNDRDHVRLALGNGLRGQVWRQFVRRPDCIYE 360
 QY 301 CRKYNTVYQYIGELLRYLCNSPQKPNDRDHVRLALGNGLRGQVWRQFVRRPDCIYE 360
 DB 301 CRKYNTVYQYIGELLRYLCNSPQKPNDRDHVRLALGNGLRGQVWRQFVRRPDCIYE 360
 QY 361 FFAATEGNIGFMNVARKVAVGVNVLQKITYDLIKYDEKDEPVDENGCYVVRPKG 420
 DB 361 FFAATEGNIGFMNVARKVAVGVNVLQKITYDLIKYDEKDEPVDENGCYVVRPKG 420
 QY 361 FFAATEGNIGFMNVARKVAVGVNVLQKITYDLIKYDEKDEPVDENGCYVVRPKG 420
 DB 361 FFAATEGNIGFMNVARKVAVGVNVLQKITYDLIKYDEKDEPVDENGCYVVRPKG 420
 QY 421 EYGLLVCKITTLTPENGVAAGAKOTEKKLRDVEFKKGLDLYNSGDLMAVDHENTFYFHDR 480
 DB 421 EYGLLVCKITTLTPENGVAAGAKOTEKKLRDVEFKKGLDLYNSGDLMAVDHENTFYFHDR 480
 QY 421 EYGLLVCKITTLTPENGVAAGAKOTEKKLRDVEFKKGLDLYNSGDLMAVDHENTFYFHDR 480
 DB 421 EYGLLVCKITTLTPENGVAAGAKOTEKKLRDVEFKKGLDLYNSGDLMAVDHENTFYFHDR 480
 QY 481 VGDTPFKMGENATTEVADTVGLVDFVQEVNNGVAVHPDHEGRIGMASIKKENHEFPKG 540
 DB 481 VGDTPFKMGENATTEVADTVGLVDFVQEVNNGVAVHPDHEGRIGMASIKKENHEFPKG 540
 QY 481 VGDTPFKMGENATTEVADTVGLVDFVQEVNNGVAVHPDHEGRIGMASIKKENHEFPKG 540
 DB 481 VGDTPFKMGENATTEVADTVGLVDFVQEVNNGVAVHPDHEGRIGMASIKKENHEFPKG 540
 QY 541 KLFQHIADYLPYSARPRFLRIODTIEITGTFKRRKMTLVEEGFNPAVAKDALYFLDDTAK 600
 DB 541 KLFQHIADYLPYSARPRFLRIODTIEITGTFKRRKMTLVEEGFNPAVAKDALYFLDDTAK 600
 QY 541 KLFQHIADYLPYSARPRFLRIODTIEITGTFKRRKMTLVEEGFNPAVAKDALYFLDDTAK 600
 DB 541 KLFQHIADYLPYSARPRFLRIODTIEITGTFKRRKMTLVEEGFNPAVAKDALYFLDDTAK 600
 QY 601 MYVPMEDIIYNAISAKTKL 620
 DB 601 MYVPMEDIIYNAISAKTKL 620
 QY 601 MYVPMEDIIYNAISAKTKL 620
 DB 601 MYVPMEDIIYNAISAKTKL 620

RESULT 3

QY VCLS_MOUSE STANDARD: PRT: 620 AA.
 ID VCLS_MOUSE
 AC 035486;

DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE VERY-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.-) (VERY-LONG-CHAIN-
 DE FATTY-ACID-COA LIGASE).
 GN FACLVL OR VLACS OR VLCS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RA Kemp S., Lu J.-F., Smith K.D.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN AMP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF033031; AAB87982.1; -
 DR InterPro: IPR000873; -
 DR Pfam: PF00501; AMP-binding: 1.
 DR PROSITE: PS00455; AMP_BINDING: 1.
 DR LIGase: Fatty acid metabolism; Peroxisome; Transmembrane.
 FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 107 127 POTENTIAL.
 FT TRANSMEM 262 282 POTENTIAL.
 SQ SEQUENCE 620 AA; 70366 MW; 77C9BDD0E3B9FFB CRC64;

Query Match 85.5%; Score 2798; DB 1; Length 620;
 Best Local Similarity 82.6%; Pred. No 5.9e-204;
 Matches 512; Conservative 52; Mismatches 56; Indels 0; Gaps 0;

QY 1 MSAIYVLAGLPLPLVNLCPYFODIGYELKVAAGRRVSGORPARTILRAFL 60
 Db 1 MLPVLYGLAGLPLPLPLFCPCYLLQDVYFQLANMARVRSYGRBPVRIILNAFL 60
 QY 61 EKASOTPKPKPLFRDELTYAAYDRSRNOYARALHDLGLRGDCAVALLMGEPAYWL 120
 Db 61 EQAKRTPKPKPLFRDELTYAAYDRSRNOYARALHDLGLRGDCAVALLMGEPAYWL 120
 QY 121 WLGVLKGCACMACLNINIRAKSLHRCGCAKYLVSPELOAVEILPSLKKDVSTY 180
 Db 121 WLGVLKGCACMACLNINIRAKSLHRCGCAKYLVSPELOAVEILPSLKKDVSTY 180
 QY 181 YVSRSTNDGIDSLFKVDYDEVSTPEDESRSEVTESTPALYITSGTGLPKRAMITHQ 240
 Db 181 YVSRSTNDGIDSLFKVDYDEVSTPEDESRSEVTESTPALYITSGTGLPKRAMITNH 240
 QY 241 RIWGTGTGTFVSGLKADVDYITLPEYHSAALLIGISVAGATTALRRKFSASQWMD 300
 Db 241 RIWGTGTGTFVSGLKADVDYITLPEYHSAALLIGISVAGATTALRRKFSASQWMD 300
 QY 301 CRKYNVTVIOYIGELRLRYLCSPOKPNDRDHKVALGNGLRGVDVWRFKRGDICTYE 360
 Db 301 CRKYNVTVIOYIGELRLRYLCSPOKPNDRDHKVALGNGLRGVDVWRFKRGDICTHYE 360
 QY 361 FYATTEGNIQGMNARVAGVAGRNVTYQKTIITLDIKIYVEKDEPVNDGTCVRRYK 420
 Db 361 FYATTEGNIQGMNARVAGVAGRNVTYQKTIITLDIKIYVEKDEPVNDGTCVRRYK 420
 QY 421 EVGLLVCKITQLPFNGYAGAKAOTEKKRLDVYKGDGLFNSGDLMLVHENVYFHDR 480
 Db 421 EVGLLVCKITQLPFNGYAGAKAOTEKKRLDVYKGDGLFNSGDLMLVHENVYFHDR 480

QY 481 VGDTPFRKGNVATTEVADTVGLVDVQEVNYYGVAPHDPHEGRIGMASIKAKENHEFDCK 540
 Db 481 VGDTPFRKGNVATTEVADTVGLVDVQEVNYYGVAPHDPHEGRIGMASIKAKENHEFNCK 540
 QY 541 KLFQHIADYLPYARPRFRIODTIEITGFRRKKTLEEGPNPVTIDALYPLDDPAK 600
 Db 541 KLFQHIADYLPYARPRFRIODTIEITGFRRKKTLEEGPNPVTIDALYPLDDPAK 600
 QY 601 MYVPMEDIYNAISAKTLK 620
 Db 601 TEVPMETINYNATIDKTLK 620

RESULT 4
 FATTY-MOUSE
 ID FATTY-MOUSE STANDARD; PRT: 646 AA.
 AC 060714;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE LONG-CHAIN FATTY ACID TRANSPORT PROTEIN (FATP).
 GN SLC27A1 OR FATP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=SWISS;
 RA MEDLINE=95042740; PubMed=7954810;
 RX Schaffer J.E., Lodish H.F.;
 RT "Expression cloning and characterization of a novel adipocyte long
 RT chain fatty acid transport protein."
 RL Cell 79:427-436(1994).
 [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98438516; PubMed=9765271;
 RA Hul T.Y., Frohner B.I., Smith A.J., Schaffer J.E., Bernlohr D.A.;
 RT "Characterization of the murine fatty acid transport protein gene and
 RT its insulin response sequence."
 RL J. Biol. Chem. 273:27420-27429(1998).
 CC -1- FUNCTION: INVOLVED IN TRANSLLOCATION OF LONG-CHAIN FATTY ACIDS
 CC ACROSS THE PLASMA MEMBRANE. MAY PLAY A PIVOTAL ROLE IN REGULATING
 CC AVAILABLE LONG-CHAIN FATTY ACID SUBSTRATES FROM EXOGENOUS SOURCES
 CC IN TISSUES UNDERGOING HIGH LEVELS OF BETA-OXIDATION OR
 CC TRIGLYCERIDE SYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PLASMA MEMBRANE.
 CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN SKELETAL MUSCLE, HEART
 CC AND FAT. LOWER LEVELS IN BRAIN, KIDNEY, LUNG AND LIVER. NO
 CC EXPRESSION IN SPLEEN OR INTESTINE.
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U15976; AAC71060.1; -
 DR EMBL: AF023258; AAC69640.1; -
 DR EMBL: AF023256; AAC69640.1; JOINED.
 DR EMBL: AF023257; AAC69640.1; JOINED.
 DR MGD: MGI:1347098; SLC27A1.
 DR InterPro: IPR000873; -
 DR Pfam: PF00501; AMP-binding: 1.
 DR PROSITE: PS00455; AMP_BINDING: 1.
 KW Glycoprotein; Lipid transport; Transmembrane.
 FT TRANSMEM 14 34 POTENTIAL.
 FT TRANSMEM 115 135 POTENTIAL.

FT TRANSMEM 140 160 POTENTIAL.
 FT TRANSMEM 169 189 POTENTIAL.
 FT TRANSMEM 293 313 POTENTIAL.
 FT TRANSMEM 337 357 POTENTIAL.
 FT CARBOHYD 330 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 518 518 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 646 AA; 71276 MW; 910B92B8BD985B4C CRC64;

Query Match 31.9%; Score 1044; DB 1; Length 646;
 Best Local Similarity 40.0%; Pred. No. 2,5e-71;
 Matches 247; Conservative 83; Mismatches 245; Indels 42; Gaps 11;

33 FLVAVVGRVRVSGY-----ORPAPTILRAFLKAROTPHKPLRDETL 79
 43 FLRIYKTKARDLFGSLVLRKRLRRRRAGDTIPCLFQAVAKRQPRRLALVVAASSGI 102
 80 --TYAODRRSNQVABALHDHGLRGDCVALLMGNPAVYVLMGLVKGACMACLN 137
 103 CMTFADLDYSNANVAN-LFROLGFAPGDVAVAFLEGRPEFVGMGLAKAGVVALLVN 161
 138 IIRAKSLHFEQCCGAVLVSPLOAAVEELPSLKDDVSIYYSRTNTGIDISFLDK 197
 162 LRREPLAFCLGTSAAKALITYGEMAAVAEVSQJGKS-----LKFCSGDLGPESILPD 216
 198 VD-----EVSTEP-----PESWRSEVTFSTALITYSTGTGLPKAMITHQRTWYGTG 247
 217 TQLLDMLAEAPTPPLAQAQKGMDRLE-----YITSGTGLPKAALVHSRYRTAA 271
 248 LRFVVS-GLKADVYITLPPYSAALLIGHCYAGATLALTRKFSASQFMDDCRKYNV 306
 272 FGHHSYSMAADVLVLDCLPLYSAGNMGVCVIGLVVLRKKFSASRFWDDCVKYN 331
 307 TVYQYIGELLRYLCSNPOKPNRBDHVKVLAIGLGLRGDWQRFVAKRGDICIYEYATE 366
 332 TVVQYIGELCRYLRLQPVADVDEQRHVRVLAVNGELRPAIMEETDQFEGVQIGEEYATE 391
 367 GNIGFMNVARKVAGVGRVNYLQKITYLDIKYDKEDEPVDEMGCVRVKRGEGVLY 426
 392 CNGSIANMGKVGSCGFNSRLITHYPRILVKNEDTMEPLDSEGLCIQCPGEPGLY 451
 427 CKTQLTP---FNGYAGAKAQTEKKRLRVFKKGLDYFNSGDLLENHENTYFHDVGD 483
 452 GQINQDDPLRFEDGYVSDSA-TNKKLAHSVFRKGSASVGLVWDELGYMYFDRSGD 510
 484 TFRWKEGNAVATPEVADTVGLVNDVGVNYGVHPDHEGRIGMASIKMKNENEDGKKLF 543
 511 TFRWKEGNAVSTEEVAVLSRLGQTDVAVYGAVPGEKAGMAAI-ADPSQLDPSNMY 569
 544 OHIADYLPSTARPEFLRDIOTTEITGTFKRRKMTVEEGFNPVAVIKDALYFLDDAKMY 603
 570 QELQKYLASARIPIFLRLPOVDITGTFRKIQTRLORESFDRQSDRLFFLIDLQGRY 629
 604 PMTEDIYMAISAKTKL 620
 630 PLDERVHARICAGDESL 646

RESULT 5
 FARP_RAT STANDARD; PRT; 646 AA.
 AC P97849;
 DT 01-NOV-1997 (rel. 35, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 30-MAY-2000 (rel. 39, Last annotation update)
 DE LONG-CHAIN FATTY ACID TRANSPORT PROTEIN (FARP).
 GN SLG27AI OR FARP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;

RP SEQUENCE FROM N.A.
 RX MEDLINE=98041635; PubMed=9375787;
 RA Schap F.G., Hamers L., van der Vusse G.J., Glatz J.F.C.;
 RL "Molecular cloning of fatty acid-transport protein cDNA from rat."
 RL Biochim. Biophys. Acta 1354:29-34(1997).
 CC -FUNCTION: INVOLVED IN TRANSLLOCATION OF LONG-CHAIN FATTY ACIDS
 CC ACROSS THE PLASMA MEMBRANE. MAY PLAY A PIVOTAL ROLE IN REGULATING
 CC AVAILABLE LONG-CHAIN FATTY ACID SUBSTRATES FROM EXOGENOUS SOURCES
 CC IN TISSUES UNDERGOING HIGH LEVELS OF BETA-OXIDATION OR
 CC TRIGLYCERIDE SYNTHESIS.
 CC -SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PLASMA MEMBRANE.
 CC -SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U89529; AAC53424.1; -
 DR InterPro: IPR000873; -
 DR Pfam: PF00501; AMP-binding; 1.
 DR PROSITE: PS00455; AMP_BINDING; 1.
 KW Glycoprotein; Lipid transport; Transmembrane.
 FT TRANSMEM 115 135 POTENTIAL.
 FT TRANSMEM 140 160 POTENTIAL.
 FT TRANSMEM 169 189 POTENTIAL.
 FT TRANSMEM 293 313 POTENTIAL.
 FT TRANSMEM 337 357 POTENTIAL.
 FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 518 518 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 646 AA; 71283 MW; C450CF14CC2EB87 CRC64;

Query Match 31.9%; Score 1042.5; DB 1; Length 646;
 Best Local Similarity 39.7%; Pred. No. 3,2e-71;
 Matches 251; Conservative 87; Mismatches 259; Indels 35; Gaps 12;

5 IYTVLAGLFLPLVNLCCPYFFQDGYFLKVAAGRRVSGQRAPRTILRAFLKAR 64
 34 YVVGSGGRFLRIYCKTARLDL-GLSVLLRVLELR-----HRRAGDTIPRIFQAVAG 87
 65 QTPKPELFRDETL--TYAODRRSNQVABALHDHGLRGDCVALLMGNPAVYVLM 122
 88 RQPRRLALVDASSGICWTFQDLDTYSNANVAN-LFOLGFAPGDVAVAFLEGRPEFGLWL 146
 123 GLVRLGAMCLNINIRAKSLHFCQCCGAKVILVSPLOAAVEELPSLKDDVSIYV 182
 147 GLAKAGVVALLVNLRREPLAFCLGTSAAKALITYGEMAAVAEVSQJGKS-----LL 201
 183 SRTSNTGIDISFLDKVD-----EVSTEP-----PESWRSEVTFSTALITYSTGTGLP 232
 202 KFCSGDLPRESVLPDQLLDMLAEAPTPPLAQAQKGMDRLE-----YITSGTGLP 256
 233 KAAMITHQRTWYGTGLFVS-GLKADVYITLPPYSAALLIGHCYAGATLALTRK 291
 257 KAAIVHVSRYRRIAPFHSYSMRANDVLDCLPLYSAGNMGVCVIGLVVLRKK 316
 292 FASQFMDDCRKYNVTVYQYIGELLRYLCSNPOKPNRBDHVKVLAIGLGLRGDWQRFV 351
 317 FASRFWDDCVKYNCTVYQYIGELLRYLCSNPOKPNRBDHVKVLAIGLGLRGDWQRFV 376
 352 RFSGDICIYEFAATEGNIGFMNVARKVAGVGRVNYLQKITYLDIKYDKEDEPVARDEN 411
 377 GFGVROGFEFGAEGCNSIANDGKVGSCGFNSRLITHYPRILVKNEDTMEPLRDSQ 436
 412 GCVRVKRGEGVGLVCKITQLTP---FNGYAGAKAQTEKKRLRVFKKGLDYFNSGDL 468

DB 437 GCLIPQGPPEGLVGOINQDPLRRFDGYSDA-TNKKIAHYERKGDASVLSGDLV 495
 QY 469 VDHENIYFHDVGTDFRKKGENVATFVADVLGVDOEVQVNYGVAVPDHEGIGMAS 528
 DB 496 MDELGMFTRDSGDVFFRKGENSESTEEVAVLSRLIGOTDVAVGVAVPGEKSGMAA 555
 QY 529 IMKENHEFDGKLLFOHIDADYLPSTARPFLRIQDTEITGTFKRRKTLVEEGENPAVI 588
 DB 556 IADPHN-QIDPMSYMOELQKVLASTAQAQPIFLRLPQVDTGTGFKIQKRLRGDFPQOT 614
 QY 589 KALYFLDDTAKMYVPMTEDINYAISAKTLK 620
 DB 615 SDRLEFLDKGGRYLPIDERVARICAGDFSL 646

RESULT 6
 PRT1_YEAST STANDARD; PRT; 623 AA.
 ID PRT1_YEAST STANDARD; PRT; 623 AA.
 AC P38225;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PROBABLE LONG-CHAIN FATTY ACID TRANSPORT PROTEIN.
 GN PAB1 OR YBR041W OR YBR041L.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=528C;
 RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,
 RA visiers S.;
 RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 RM [2]
 RP CHARACTERIZATION.
 RC STRAIN=M303A;
 RX MEDLINE=97235610; PubMed=9079682;
 RA Faergeman N.J., Dirusso C.C., Elberger A., Knudsen J., Black P.N.;
 RT "Disruption of the Saccharomyces cerevisiae homologue to the murine
 RT fatty acid transport protein impairs uptake and growth on long-chain
 RT fatty acids." 272:8531-8538(1997).
 RL J. Biol. Chem. 272:8531-8538(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN LONG-CHAIN FATTY ACIDS UPTAKE, AND
 CC THIS MAY PLAY A PIVOTAL ROLE IN REGULATING THEIR ACCESSIBILITY
 CC PRIOR TO METABOLIC UTILIZATION. MAY PLAY AN IMPORTANT ROLE IN
 CC UPTAKE OF THESE HYDROPHOBIC COMPOUNDS UNDER CONDITIONS WHERE FATTY
 CC ACID SYNTHESIS IS COMPROMISED.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT
 CC BINDING OF AMP TO THEIR SUBSTRATE.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z35910; CA84983.1; -
 CC PIR: S45899; S45899.
 CC SGD: S0000245; FATT1.
 CC InterPro: IPR000873; -
 CC Pfam: PF00501; AMP-binding; 1.
 CC PROSITE: PS00455; AMP BINDING; 1.
 KW Lipid transport; Transmembrane.
 FT TRANSMEM 6 26 POTENTIAL.
 FT TRANSMEM 54 71 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 FT TRANSMEM 293 313 POTENTIAL.
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 623 AA; 71697 MW; 0AE02700B60C8CFE CRC64;
 Query Match 22.8%; Score 744.5; DB 1; Length 623;
 Best Local Similarity 32.8%; Pred. No. 1e-48; Indels 59; Gaps 16;
 Matches 191; Conservative 96; Mismatches 236;
 QY 24 PYFQDIDGYFLKVAAGR-----FVRSYGQRPPARTILRAFLERAKQTPKPF 72
 DB 58 PYFLKSY--FCYILDVARRHFQNMWYLEFKQVQNGDHL-AISTRRMAEKG----- 106
 QY 73 LFRDELITYAQNDRSRQVAPALHDHLLGKQDCCVALLMGNEPAYVWLGLVLCGAMA 132
 DB 107 -FOLEPTTYIETNYIVRLSHILHFDYNNQAGDYVAIDCTNKLDFVFLWLSLNIIGAIPA 165
 QY 133 CLANYINRAKSLHFOCCGAKVLLVSPDLQAAYEELPSLKK--DDVSITYYSRTSNTDG 190
 DB 166 FLNNTKGTPLVSHKLSNITQVFIIDPDASNPIRESEELKNAIPVKLNYLEEQ----- 220
 QY 191 IDSLDKVDVSTEEPIPEWSRSE-----VTESTPALYIYTSQPTGLPKAMITHQRTM 243
 DB 221 ----DLMHLLNSQSPSEFLQDNVETPLGLTDFKPSMLIYTSQPTGLPKASIMSRKS 275
 QY 244 YGTGL-TFVSGKLADVDYITTFPFYSAALLIGHCYAGATLARTFSAQFDDCR 302
 DB 276 VGCQVGHVHMTNESTVEFAMPLFHSHTALLGACALISHGCLASHFSAFTEFKQVY 335
 QY 303 KYNVTYQYIGELLYRLCNSPQKPNDRHKVRLANGLRGDVGWRFVRFQDICYEY 362
 DB 336 LTGATHQYQVEGCRILHLPISKEKMKVKAAYNGRLPDWQFRFRFNEYGEY 395
 QY 363 AATGNTGFMNVR---KYGAVRVNTLQKITYY--LTKDVEKDEPV-RDENCYVR 416
 DB 396 AATEAPFATTTFOKQDFGIGACRNYGTIIQWFLSFQOTLVKMDPNDVSYRNSKFCV 455
 QY 417 VPFQVGLVLCXI-----TQTPENGYAGAKAQEKKLLDVPFKGGLYNSGDLMTVDE 472
 DB 456 APVGEFGEKMLFFPKKPEPSTFQGLVNAKKEKSVVDVFRFGAWRCGDLKADXY 515
 QY 473 NFIYFHDVGDPRMKGENVATTEVAD--TVGLVDFQVQVNYGVAVPDHEGRIGMASIK 530
 DB 516 GLWYFLDRMGDTFRKMSSENVSTTEVEDQLTASKEQYAVIVGVIVPEBGRAGAVIK 575
 QY 531 MKENH-EFDCK-KLFGHID--VLPSYARPRIRIDTIEIT 568
 DB 576 LTDNSLDTAKTKRLKLNDSRLNLPYAMPLEFKVDEIKMT 617

RESULT 7
 CATC_ECOLI STANDARD; PRT; 522 AA.
 ID CATC_ECOLI STANDARD; PRT; 522 AA.
 AC P31552;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PROBABLE COTONOBETAINE/CARNITINE-COA LIGASE (EC 6.3.2.-).
 GN CATC.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=O4 K74;
 RX MEDLINE=95115548; PubMed=7815937;
 RA Eichler K., Bourgis F., Buchet A., Kieber H.-P.,
 RA Mandrand-Berthelot M.-A.;
 RT "Molecular characterization of the cat operon necessary for carnitine
 RT metabolism in Escherichia coli.";
 RL Mol. Microbiol. 13:775-786(1994).
 RN [2]

Dd	253	VYTAAPLPIEAFLLT---	NCLFIEIGSGONLLITPNDRIPGLVKELAKFYFAITGVNTL	309
Oy	316	LRYLCSPKPRNDRHVKAFLALGNL-	RGDWVRQVKKRFGSDICITYEAATE-----	366
Dd	310	FNALNNKEFOOLDPSSLSHLSAGGHPVOQVAERWKVLTGTQ-	LLEGGLGLECAPLVASV	368
Oy	367	-----GNIGFMNAYARKVAGAVGRVNYLQKKIITIYYDLIKDYDEKEDPRNGCYVAV	417	
Dd	369	NPYDIDHYHSGSIGL-----	PVPSTEAKLV-----DD- ---DN-----EV	399
Oy	418	PKGEGLLYCKITQLTPFNMGYGAAQAQTCKKLRLDYFKKGIDLXFNSGDILAMDHENFIYE	477	
Dd	400	PGGPRELCLKRPQV--MLGIWORPDAD----	EIIKNG--WLHTGDIAVNDEREGFLRI	450
Oy	478	HDRVDTFRMKGENVATTEVDATVGLVDFVEQENYVGYVDPDEHRIGMASIKKENHEFF	537	
Dd	451	YVRKKDMILVSSENVYPNIIEDVVQHREVGLEAVALG--	VPSGSSEANAKIRIVVKKDPSL	508
Oy	538	DGKKLFQHLADVLPSYARREFLRIDTLTIETGFHKRKTTLYEE	581	
Dd	509	TESLVTCTCRRLTGKYVKKLVEFDELKPSNVGKILRELRDE	552	
RESULT	9			
AC	4CL_VANPL	STANDARD:	PRT;	553 AA.
AC	024540:			
Dt	01-OCT-2000	(Rel. 40, Created)		
Dt	01-OCT-2000	(Rel. 40, Last sequence update)		
Dt	01-OCT-2000	(Rel. 40, Last annotation update)		
De	-4-COUMARATE--COA LIGASE	(EC 6.2.1.12) (CCL) (-4-COUMAROYL--COA SYNTHASE).		
Gn	4CL			
Os	Vanilla planifolia (Vanilla).			
Oc	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
Oc	Magnoliopsida; Liliopsida; Asparagales; Orchidaceae; Vanilla.			
Ox	NCBI_TaxID=51239;			
Rn	[1]			
Rp	SEQUENCE FROM N.A.			
Ra	Brodellus P.; Xue Z.T.;			
Rt	"Isolation and characterization of a cDNA from cell suspension cultures of Vanilla planifolia encoding 4-coumarate: coenzyme A ligase.";			
Rt	Plant Physiol., Biochem. 35:497-506(1997).			
Cc	-1- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE + 4-COUMAROYL--COA.			
Cc	-1- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.			
Cc	-1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT COVALENT BINDING OF AMP TO THEIR SUBSTRATE.			
Cc	-----			
Cc	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
Cc	-----			
Dr	EMBL; X75542; CAAS3230.1; -.			
Dr	HSSP; P08659; ILCI.			
Dr	Mendel; 24370; Vanpl;1179;24370.			
Dr	InterPro; IPR000873; -.			
Dr	Pfam; PF00501; AMP-Binding; 1.			
Dr	PROSITE; PS00445; AMP-BINDING; 1.			
Kw	Ligase; Phenylpropanoid metabolism.			
Sq	SEQUENCE 553 AA; 60095 MW; 9A2D250BC84EALCB CRC64;			

[illegible]

DT 01-OCT-2000 (rel. 40, Last annotation update)
 DE 4-COMMATE-1-COA LIGASE 1 (EC 6.2.1.12) (4CL 1) (4-COMMATE-1-COA
 GN 4CL1 OR 4CL)
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
 OC Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. JAPONICA;
 RX MEDLINE=91045096; PubMed=2235510;
 RA Zhao Y., Kung S.D., Dube S.K.;
 RT "Nucleotide sequence of rice 4-coumarate:COA ligase gene, 4-CL.1";
 RL Nucleic Acids Res. 18:6144-6144(1990).
 CC -1- CATALYTIC ACTIVITY: ATP + 4-COMMATE + COA = AMP + DIPHOSPHATE +
 CC 4-COMMATE-1-COA.
 CC -1- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID
 CC METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.
 CC -1- INDUCTION: TRANSCRIPTIONALLY ACTIVATED BY BOTH FUNGAL ELICITOR AND
 CC ULTRAVIOLET IRRADIATION.
 CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
 CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X52623; CA36850.1; -
 DR PIR; J00311; J00311.
 DR HSSP; P08659; ILCI.
 DR InterPro; IPR000873; -
 DR Pfam; PF00501; AMP-binding; 1.
 DR PROSITE; PS00455; AMP-BINDING; 1.
 KM Ligase; Phenylpropanoid metabolism; Multigene family.
 SQ SEQUENCE 563 AA; 60841 MW; 35B329401C6B44B9 CMC64;

Query Match 8.3%; Score 270.5; DB 1; Length 563;
 Best Local Similarity 24.1%; Pred. No. 6,1e-13;
 Matches 125; Conservative 82; Mismatches 234; Indels 77; Gaps 21;
 QY 79 LTYAVDRSRNGVAYAL-HDHILGQSGCVALLMNGENAPAYWMLGLYKLCACACLTNTN 137
 DB 65 LTYAVDRSRNGVAYAL-HDHILGQSGCVALLMNGENAPAYWMLGLYKLCACACLTNTN 137
 QY 138 IRASLILHCFQCCAKVLLVSPQLAAVEELPSLKKDDIYVYRSNTDGIIDSLDK 197
 DB 125 STPHIESQLAAGATVIT---ESMAADKLPSHSHALTVLID--ERDDCLHFDD 178
 QY 138 V---DEVSTEPPESEWSEVTFSPALYI--YTSGTGLPRAAMITHQRIWYGTGLFVS 252
 DB 179 LMSDEAS--PLAGDEDEKVFDDVVALPYSGTGLPGVMVLTIRSL-----STSYA 231
 QY 253 -----GLKADVYITLPPYH---SAALLIGHGIVAGTLALTKTSAGQFW 298
 DB 232 QOVGGENNIGLHAGDYLCLAPMFHYSLTIMCGLR---VGAIVVRRRDLAAM 287
 QY 299 DDCRRYNTVTOYIGELLRYLCSPOKENDRHVRLALGNG--LRGDVWRFKRGDI 356
 DB 288 DLVEHRYVTIAPLYPIYVAVAKSBAARLSSVBMVLSGAAPMGKIDIAFAKALPGA 347
 QY 357 CIYEYATTEGN-----IGFMNARKY--GAVGR-VNYLQKTIITDLYDKVEKDEPV 407
 DB 348 VLGQGYGTEAGPYLSMCLAFKPEKVGKSGAGCTVVVNNALKTIIDPTGR---SLGRNL 404
 QY 408 RDENGYCVRVKGEVGLVCKITQITLTPNGYAGAKQTEKKKLDVPRKKGDLVENSDDL 467

DB 405 RGE--ICIR-----GGQIMK-----GYNNPEATK-----NTIDAEG--WHLHGIDG 442
 QY 468 MYDENFTYFDRVGDYFRKNGENVATTEVADYGLVDFVQEVNNGVHPDHEGRIGMA 527
 DB 443 YVDDDELFYDRUKETIKYGFQVAPAELEALLNTHPSIDADAVALKPF-----GEIPVA 498
 QY 528 SIKKENHEFGDKTLFQIHADYLPSPVAPRFRIRQDTI 565
 DB 499 FVATGESSELSDDYKQVAKETIYIKIREFFVDKI 536
 RESULT 12
 SRFL_BACSU STANDARD; PRT; 3587 AA.
 ID SRFL_BACSU
 AC P27206;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SURFACTIN SYNTHETASE SUBUNIT 1.
 GN SRFA OR SRFA1 OR SRFA.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=93181186; PubMed=8441623;
 RA Fuma S., Fujishima Y., Corbelli N., D'Souza C., Nakano M.M.,
 RA Zuber P., Yamane K.;
 RT "Nucleotide sequence of 5' portion of srfa that contains the region
 RT required for competence establishment in Bacillus subtilis.";
 RL Nucleic Acids Res. 21:93-97(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / JH642;
 RX MEDLINE=93360813; PubMed=8355609;
 RA Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
 RA Venema G., van Sinderen D.;
 RT "Sequence and analysis of the genetic locus responsible for surfactin
 RT synthesis in Bacillus subtilis.";
 RL Mol. Microbiol. 8:821-831(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=97124189; PubMed=8969502;
 RA Yamane K., Kumano M., Kurita K.;
 RT "The 25 degrees-36 degrees region of the Bacillus subtilis
 RT chromosome: determination of the sequence of a 146 kb segment and
 RT identification of 113 genes.";
 RL Microbiology 142:3047-3056(1996).
 RN [4]
 RP SEQUENCE OF 1-460 FROM N.A.
 RX MEDLINE=91154134; PubMed=1847909;
 RA Nakano M.M., Magnuson R., Myers A.M., Curry J., Grossman A.D.,
 RA Zuber P.;
 RT "srfa is an operon required for surfactin production, competence
 RT development, and efficient sporulation in Bacillus subtilis.";
 RL J. Bacteriol. 173:1770-1776(1991).
 RN [5]
 RP SEQUENCE OF 1-38 FROM N.A.
 RX MEDLINE=91358326; PubMed=1715856;
 RA Nakano M.M., Xia L., Zuber P.;
 RT "Transcription initiation region of the srfa operon, which is
 RT controlled by the comp-coma signal transduction system in Bacillus
 RT subtilis.";
 RL J. Bacteriol. 173:5487-5493(1991).
 RN [6]
 RP SEQUENCE OF 1-64 FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95219080; PubMed=7704255;
 RA Fujishima Y., Yamane K.;
 RT "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)

RT of srfa of the Bacillus subtilis chromosome.;

RL Microbiology 141:277-279(1995).

CC -1- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO

CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.

CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.

CC -1- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETHEINES.

CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE

CC ANTIBIOTIC SUBFACTIN.

CC -1- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT

CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.

CC -----

CC This swiss-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; D13262; BAA02522.1; -

DR EMBL; X70356; CAA49816.1; -

DR EMBL; D50453; BAA08982.1; -

DR EMBL; M59939; AAA22815.1; -

DR EMBL; M64702; AAA22816.1; -

DR EMBL; D30762; BAA21034.1; -

DR EMBL; Z99105; CAB12142.1; -

DR PIR; S35517; S35517.

DR HSSP; P14687; LAMU.

DR Subtilist; BG10168; srfaA.

DR Interpro; IPR000285; -

DR Interpro; IPR000873; -

DR Interpro; IPR001242; -

DR Pfam; PF00501; AMP-binding; 3.

DR Pfam; PF00668; DUF4; 4.

DR Pfam; PF00550; pp-binding; 3.

DR PRINTS; PR00154; AMPBINDING.

DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 3.

DR PROSITE; PS00455; AMP_BINDING; 3.

DR PROSITE; PS00075; ACP_DOMAIN; 3.

KW Ligase; Antibiotic biosynthesis; Phosphopantetheine; Sporulation;

KW Multifunctional enzyme; Repeat.

FT REPEAT ? 2086

FT REPEAT ? 3114

FT DOMAIN 976 1043

FT DOMAIN 2015 2082

FT DOMAIN 3043 3109

FT BINDING 1006 1006

FT BINDING 2045 2045

FT BINDING 3073 3073

FT BINDING 146 146

FT CONFLICT 151 151

FT CONFLICT 165 165

FT CONFLICT 281 281

FT CONFLICT 460 460

FT CONFLICT 540 540

FT CONFLICT 562 562

FT CONFLICT 639 640

FT CONFLICT 644 644

FT CONFLICT 647 647

FT CONFLICT 649 649

FT CONFLICT 1026 1026

FT CONFLICT 1065 1116

FT CONFLICT 1132 1133

FT CONFLICT 1164 1164

FT CONFLICT 1458 1458

FT CONFLICT 1850 1850

FT CONFLICT 1894 1894

FT CONFLICT 1911 1911

FT CONFLICT 1973 1980

DS -> VC (IN REF. 2 AND 3).

R -> A (IN REF. 2 AND 3).

V -> L (IN REF. 2 AND 3).

S -> T (IN REF. 2 AND 3).

D -> T (IN REF. 2 AND 3).

R -> A (IN REF. 2 AND 3).

L -> Q (IN REF. 2 AND 3).

L -> O (IN REF. 2 AND 3).

SLAQRRTYVSGFAGVGYNNPAALLEGPLDIOKTERAF

OGLIRRHSLR -> HHMSSEHSSASSRMSAIIHASSSN

SEGFYSKRGARISGINPPRYIE (IN REF. 2 AND

3).

FT CONFLICT 2052 2052 L -> P (IN REF. 2 AND 3).

FT CONFLICT 2218 2218 H -> Q (IN REF. 2 AND 3).

FT CONFLICT 2265 2266 QQ -> HR (IN REF. 2 AND 3).

FT CONFLICT 2291 2291 V -> L (IN REF. 2 AND 3).

FT CONFLICT 2349 2349 S -> C (IN REF. 2 AND 3).

FT CONFLICT 2428 2428 R -> S (IN REF. 2 AND 3).

FT CONFLICT 2611 2612 RC -> AV (IN REF. 2 AND 3).

FT CONFLICT 2685 2685 E -> ENPE (IN REF. 2 AND 3).

FT CONFLICT 2755 2755 T -> S (IN REF. 2 AND 3).

FT CONFLICT 2897 2897 SP -> TA (IN REF. 2 AND 3).

FT CONFLICT 3025 3025 P -> N (IN REF. 2 AND 3).

FT CONFLICT 3096 3096 F -> N (IN REF. 2 AND 3).

FT CONFLICT 3271 3271 A -> S (IN REF. 2 AND 3).

FT CONFLICT 3316 3316 R -> S (IN REF. 2 AND 3).

FT CONFLICT 3451 3451 Y -> S (IN REF. 2 AND 3).

FT CONFLICT 3483 3484 DE -> HQ (IN REF. 2 AND 3).

FT CONFLICT 3487 3490 DAGL -> HPEF (IN REF. 2 AND 3).

FT CONFLICT 3493 3493 R -> T (IN REF. 2 AND 3).

FT CONFLICT 3495 3495 E -> H (IN REF. 2 AND 3).

FT CONFLICT 3499 3500 GQ -> PH (IN REF. 2 AND 3).

FT CONFLICT 3507 3507 R -> E (IN REF. 2 AND 3).

FT SEQUENCE 3587 AA; 402424 MW; A1E2DABFF93JED3A CRC64;

Query Match 8.1%; Score 266.5; DB 1; Length 3587;

Best Local Similarity 23.1%; Pred. No. 1.5e-11;

Matches 126; Conservative 84; Mismatches 249; Indels 87; Gaps 16;

QY 53 RTIIRAPLEKARQTPPKPELLFRDETITVAQVDRSRQVRAALHDHGLRGSGCVALLMG 112

DB 1505 KTIYOLFEEQANPPDFTALQYBESLTYRELENNARAKLRGLIIS-LGAEGSTAAVLCE 1563

QY 113 NEPAVYVWLGLVLCGAMACLNINIRAKSLHFCQCGAKVLLVSPELQAAVEETLPST 172

DB 1564 RSMQMIYSLALVLEKSGSAYVPIDPEHDIQRMQHFRSGAKVLLTQKKLALEE----- 1618

QY 173 KKDVSIIYYSRTSNTGICIDSLDKVDVST-----PIPEWSRSVYTSFALYIYS 226

DB 1619 -----AEFGVYVLADEESYHADARNALPLDSAMANLT-----YYS 1657

QY 227 GTTGPRKAMITTHORIVWGTGLFVSGLKADVDYIYLPFYSAALLIGIGCVAGATL 286

DB 1658 GTTGTPKGNITTHANILRYKETYLYSITTEDIT-LGLSNVYDAPFDFGSLNAGKL 1716

QY 287 AL---RTKFSASQFWDCKRKYNVTIOYIGELLRYL-----CNSPKPNDRDHKVRLAL 337

DB 1717 VLIPKEFVLDGARLSRVIERENISILMITTALFHLVDLPACLS-----TLRKIM 1767

QY 338 GNGLRGV--WRQVKKRPGDICIYEFAATGNGINGMNARKVAGARVNYLQKKIITYD 395

DB 1768 FGEERASVEHVRKALQYVKGKLLHMYGPSESTV-FATY-----HPVDELEHTLSVP 1819

QY 396 LIKYDVEKDEVPBRENCGVAPRKGEGVLVCKITQTLPPNGYAGAKQTEKKLRDVF 455

DB 1820 IGRVPSNTEVYIIDRTGIV--QPRAGINGELCVSGELV--KGYNNRRELTEKFPVPHPT 1875

QY 456 KGDLYFNSGILLVADHENEYFHDVGDTEFRKGENVATTEVADTVGLVDFVQEVNYYG 515

DB 1876 SGERNYKTKGDLARLMPNDIEFIRIDHOKVIRGORIETLEIEHQLOTHRVQESVYLA 1935

QY 516 HVPDHEGR-----IGMASIKKENHEPDGKKLFQHLADYLPVARRPRLRIDTIE 566

DB 1936 ---DQGGDKLLCAYYVGBDISSQEMRE-----HAAKDLPAVMVLRCLSKWTLP 1983

QY 567 ITGTGF 572

DB 1984 LTGNK 1989

RESULT 13

4CL1_PETCR

ID 4CL1_PETCR

AC P14912; STANDARD; PRT; 544 AA.

[illegible][illegible]

DB 454 ALLVAHPSTADAAV---VPQKHEEAGEVPVAFVYKSSSEISQETKEFVAKOVIFPKKH 509
QY 544 -QHADYLPSTYARPRFLR 560
DB 510 RYFVDAIPKSPSGKILR 527

Search completed: July 16, 2001, 18:15:11
Job time: 187 sec

2